Accounting for female space sharing in St. Kilda Soay sheep (Ovis aries) results in little change in heritability estimates

Title: Accounting for female space sharing in St. Kilda Soay sheep (*Ovis aries*) results in little change in heritability estimates

Running title: Does philopatry bias heritability estimates?

Charlotte. E. Regan - Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK - C.E.Regan@sms.ed.ac.uk

Institute of Evolutionary Biology
School of Biological Sciences
University of Edinburgh
Ashworth Laboratories
Charlotte Auerbach Road
Edinburgh
EH9 3FL

Telephone: +44 (0)131 650 5440

Jill. G. Pilkington - Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK - J.Pilkington@ed.ac.uk

Camillo Bérénos - Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK - C.Berenos@ed.ac.uk

Josephine. M. Pemberton - Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK - J.Pemberton@ed.ac.uk

Per. T. Smiseth - Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK - Per. T.Smiseth@ed.ac.uk

Alastair. J. Wilson - Centre for Ecology and Conservation, University of Exeter, Penryn campus, Cornwall, UK - A.Wilson@exeter.ac.uk

Acknowledgements

We are grateful to the National Trust for Scotland and Scottish Natural Heritage for permission to work on St Kilda, and QinetiQ and Eurest for logistics and other support on the island. We further thank all project members and many volunteers who have made great contributions to field work on the island, and all those who have contributed to keeping the project going over many years, including T. Clutton-Brock, M. Crawley, S. Albon, T. Coulson, L. Kruuk, A. Wilson, and D. Nussey. We are also grateful to Jarrod Hadfield for helpful discussion and comments that greatly improved the analysis. The Soay Sheep Project is supported by grants from the UK Natural Environment Research Council, whilst CER is supported by a BBSRC PhD studentship.

Abstract

When estimating heritability in free-living populations, it is common practice to account for common environment effects, because of their potential to generate phenotypic covariance among relatives thereby biasing heritability estimates. In quantitative genetic studies of natural populations, however, philopatry, which results in relatives being clustered in space, is rarely accounted for. The two studies to have done so suggest

absolute declines in heritability estimates of up to 43% when accounting for space sharing by relatives. However, due to methodological limitations these estimates may not be representative. We used data from the St. Kilda Soay sheep population to estimate heritabilities with and without accounting for space sharing for five traits for which there is evidence for additive genetic variance (birth weight, birth date, lamb August weight, and female post mortem jaw and metacarpal length). We accounted for space sharing by related females by separately incorporating spatial autocorrelation, and a home range similarity matrix. Although these terms accounted for up to 17% of the variance in these traits, heritability estimates were only reduced by up to 7%. Our results suggest that the bias caused by not accounting for space sharing may be lower than previously thought. This suggests that philopatry does not inevitably lead to a large bias if space sharing by relatives is not accounted for. We hope our work stimulates researchers to model shared space when relatives in their study population share space, as doing so will enable us to better understand when bias may be of particular concern.

Key words: philopatry, additive genetic variance, 'animal model', home range, common environment

Introduction

Animal breeders and evolutionary biologists often want to estimate a trait's evolutionary potential. To do this, we estimate genetic components of variance for, and covariance between, traits of interest. In the simplest univariate case, studies typically focus on the additive genetic variance (V_A) and narrow-sense heritability (h^2, the ratio of V_A to phenotypic variance). Through quantitative genetic models, these parameters can be estimated for quantitative traits using data on the phenotypic similarities of individuals of known relatedness (Falconer & Mackay, 1996; Lynch & Walsh, 1998). The development of the 'animal model', a type of mixed effects model, has greatly advanced the application of quantitative genetic analysis to wild populations. This is because the animal model uses information from individuals of varying degrees of relatedness, can cope with missing links in the complex pedigrees so typical of wild populations, and is flexible enough to incorporate known or hypothesised non-genetic influences on the phenotype (Wilson et al., 2010). Non-genetic influences on the phenotype can come from a variety of sources. In general, if environmental conditions affect phenotypes, then individuals that share a similar environment will have similar phenotypes (but note that there are exceptions: for example, sibling competition can generate greater within-brood variation in growth and survival (Nilsson & Svensson, 1996). Environmental effects that are shared by groups of individuals are referred to as 'common environment' effects (Falconer & Mackay, 1996), and these effects generate increased phenotypic similarity. In experimental studies, it is standard practice to break up the association between

genes and the local environment by distributing families across, for example, cages or tanks. Such an approach is not generally feasible in the wild, and therefore statistical techniques are used to account for common environment effects (e.g. birth year or habitat type) by including them as fixed or random effects (e.g. McCleery et al. 2004; Vergara et al. 2015). Cross-fostering has however been used to separate out environmentally derived similarity from that due to shared genes in some studies of birds (Hadfield et al., 2006), and mammals (McAdam et al., 2002). A combination of cross-fostering and the animal model is the best way to avoid bias in genetic parameters when common environment effects are strong (Kruuk & Hadfield, 2007), however cross-fostering is not feasible in all systems used to study quantitative genetics in the wild, for example the ungulates.

Relatives are often clustered in time and/or space, and therefore often share environments as well as genes. Where this is the case, common environment effects can be particularly problematic, resulting in biased heritability estimates because we make the mistake of assuming that their similarity is due to shared genes alone (Kruuk & Hadfield, 2007). For example, maternal effects result in offspring born to the same mother being more similar to one another than offspring from different mothers (Falconer & Mackay, 1996). Therefore, failing to account for maternal effects can upwardly bias estimates of V_A, and consequently h^2 (Wilson et al., 2005). As a result, it is now routine to account for maternal effects when conducting quantitative genetic analysis. Other sources of common environment effects however have received less attention in quantitative genetic analyses of natural populations. For example, we tend to neglect the fact that relatives can experience similar environments even into adulthood, as a result of natal philopatry (e.g. Rossiter et al. (2002)). If this is the case, and the environment is spatially heterogeneous, then we might expect relatives to be more phenotypically similar, because they experience more similar environments. In other words, the value of a trait expressed in an individual may be related to the trait values of individuals at nearby locations, a phenomenon known as spatial autocorrelation (SAC) (Cliff & Ord, 1981; Olalla-Tárraga et al., 2007; Ng et al., 2013). As with maternal effects, failing to account for this type of common environment effect has the potential to bias estimates of V_A and h^2. The potential for SAC to be a source of bias in genetic parameter estimates has been of concern to plant breeders for some time (Cullis & Gleeson, 1989, 1991; Magnussen, 1993; Qiao et al., 2000), particularly in the case of forestry and agricultural variety trials (Dutkowski et al., 2002). Traditionally, experimental design was used to combat this problem, but was often unsuccessful because of the variability in the patterns and scale of spatial variation, resulting from differences in the underlying gradients, ranging from soil and microclimatic effects, to cultural and measurement effects (Dutkowski et al., 2002). Statistical techniques to explicitly model SAC in analyses primarily aimed at estimating genetic parameters have therefore become more popular (Dutkowski et al., 2002). Though the addition of an SAC term generally results in model improvement, the effect of doing so on the genetic variance

is variable, with both increases and decreases reported in the plant breeding and forestry literature (Silva et al., 2001; Dutkowski et al., 2002, 2006; Banerjee et al., 2010).

Although studies on plants illustrate that accounting for spatial sources of similarity can be important in deriving accurate heritability estimates, to our knowledge there have only been two studies that have considered space sharing by relatives beyond the immediate natal environments when conducting quantitative genetic studies on wild animal populations (but see Heckerman et al. 2016 for a recent human study). Firstly, a study of laying date and clutch size in the Wytham wood great tit (*Parus major*) population, found that accounting for SAC resulted in an absolute decrease of 25% (from 40% to 15%) in the estimated heritability of laying date, though no such trend was evident for clutch size (Van Der Jeugd & McCleery, 2002). Secondly, a study on the red deer (*Cervus elaphus*) population on the Scottish island of Rum found evidence consistent with space sharing being an important source of bias in heritability estimates (Stopher et al., 2012). In this study, the change in the estimated heritability varied substantially, from an absolute change of 43% (from 44% to <1%) in the case of spring home range size to only around 4% for lifetime breeding success (from 4% to <1%) (Stopher et al., 2012). Although these studies have greatly advanced our understanding of how failure to account for spatial structure in wild populations may bias heritability estimates, there is a need to build on these works, using improved methodologies to understand how heritability estimates are affected when space sharing by relatives is not or cannot be accounted for. Firstly, we need to continue to develop methods to account for space sharing within the animal model, given that heritability estimates derived from out-dated techniques, such as parent-offspring regression are less accurate (Kruuk, 2004; Akesson et al., 2008). Of the two studies mentioned above, only the one by Stopher et al. (2012) used the animal model approach, while Van Der Jeugd & McCleery (2002) conducted parent-offspring regressions for three groups of individuals whose nestboxes were separated by varying distances. The extension of this approach to additional traits, populations and species will be necessary before there can be any general conclusions about the effect of accounting for space sharing by relatives on heritability estimates. Secondly, we need to make use of the sophisticated methods available to quantify individual space use, such as utilisation distributions (UDs - a relative frequency distribution describing the probability of an individual occurring at a particular location at a specific point in time) (Worton, 1989; Kie et al., 2010). Such techniques are however sensitive to the availability of location data (Seaman et al., 1999; Blundell et al., 2001), and the inclusion of individuals with few observations may have influenced the results of Stopher et al. (2012) through the under- or over-estimation of space use similarity. Thirdly, trait choice is likely to be important when drawing conclusions about the severity of the bias in heritability estimates as a result of not accounting for the space sharing of relatives. For example, Stopher et al. (2012) found large decreases in heritability estimates for two home range size (spring and rut) traits when accounting for the space sharing of related animals, leading them to conclude

that heritability estimates can decrease dramatically when space sharing is accounted for. Given that they are spatial metrics, the home range size traits were very likely to have a spatially autocorrelated component. They were therefore useful to demonstrate that similarity in shared space can appear as similarity due to shared genes, providing an example of the potential severity of the bias when failing to account for space sharing by relatives. However, there is to our knowledge little evidence to suggest that such traits have a heritable basis, particularly in mammals where home range size has been shown to vary with a wide variety of factors (van Beest et al., 2011). The results for these traits are therefore unlikely to prove representative of the degree of bias in quantitative genetic parameters. There is a need to build on the study by Stopher et al. (2012), examining a wider range of traits, and focusing in particular on those that, based on previous research, are believed to be heritable. Indeed, although it is sensible to account for all suspected common environment effects when aiming to accurately estimate heritability, this may not always be possible given data limitations. Therefore studies are needed to better establish the likely extent of the bias in traits as a result of not accounting for such common environment effects.

Quantitative genetic analyses of wild populations are continuing to grow in popularity (Kruuk et al., 2008). This means it is essential to expand our understanding of potential biases in heritability estimates due to space sharing by related individuals, making use of the rapidly developing methodologies. The St. Kilda Soay sheep (*Ovis aries* Linnaeus, 1758) population is an ideal system for doing this. Firstly, females are philopatric, with relatedness increasing with home range proximity (Coltman et al., 2003). As a result, any phenotypic similarity between related females may be partially due to common environment effects resulting from space sharing. Secondly, there is spatial heterogeneity in the environment. Forage availability and quality varies markedly across the study area (Regan et al., 2015), with the highest quality grazing found in the previously cultivated meadows, and increasing density of low palatability species such as *Calluna vulgaris* as elevation increases (Coulson et al., 1999). Thirdly, because the population has been studied intensively for 30 years, we have sufficient data to quantify individual ranging behaviour and relatedness, making it possible to run animal models which include information on individual space use. Indeed, this population has been the focus of quantitative genetic analysis for many years, providing an ideal platform for expanding on these modelling approaches. Furthermore, in contrast to many other long term studies of natural populations, a genomic relatedness matrix is available in place of a traditional pedigree. The use of this matrix has been recently shown to give more precise quantitative genetic estimates (Bérénos et al., 2014).

We aimed to understand how accounting for space sharing by related females affected our estimates of V_A or V_{MG} (maternal genetic effects) and h^2_T (the total heritability - accounting for additive and maternal genetic effects) or h^2 (the narrow-sense heritability) for five traits that are, based on previous research, believed to have a heritable basis (birth weight, birth date, lamb August weight, adult jaw length and

adult metacarpal length). We predicted that individuals which were similar in their space use would be more similar in their phenotype (or the phenotype of their lambs), and that this would be particularly pronounced for birth weight, birth date and August weight, because these traits are closely tied to resource availability. Consequently, we also expected considerable bias in heritability estimates when space sharing was not accounted for. We provide only the second study to look at the effect of space sharing on estimates of heritability. Using improved methodologies we show that heritability estimates may be less affected by this source of common environment effect than previously thought.

Methods

Study population and data collection

The data used in this paper come from the Soay sheep population on the island of Hirta in the St. Kilda archipelago, Scotland (57°49° N 08°34° W). This population has been unmanaged since its introduction from the neighbouring island of Soay in 1932 (Clutton-Brock et al., 2004), and Hirta is now home to between 700 and 2300 Soay sheep, depending on variation in mortality between years. Sheep residing in the Village Bay area of Hirta make up approximately one third of the total island population, and have been intensively studied since 1985 (Clutton-Brock et al., 2004).

The majority of lambs are ear-tagged within the first few days of life, making individuals uniquely identifiable. The mortality status of animals is tracked through regular censuses and mortality checks, with the census data also providing information on individual space use. Each August approximately two thirds of the Village Bay population are caught, at which time each individual is weighed. Because mortality is tracked closely, we are also able to take post mortem trait measurements from many animals, including jaw and metacarpal length. We selected three early life traits, birth weight, birth date and lamb August weight (all treated as a trait of the lamb), and two adult traits, female post mortem jaw length, and female post mortem metacarpal length. These traits were selected because they had previously been the focus of quantitative genetic study, and because of their potential link with resource availability. See Table 1 for heritability estimates for these traits from previous studies. For the adult traits, we incorporated information on the space sharing of all females with post mortem jaw and metacarpal length measurements. For the early life traits we used information on the space sharing of their mothers because at the point of measurement lambs have not developed their own home range. There are strong maternal genetic effects in all three early life traits (Wilson et al., 2005; Bérénos et al., 2014), and we were therefore interested in the change in this term when accounting for the space sharing of related mothers. There is no evidence for significant maternal genetic

effects for either of the adult traits and therefore we did not estimate them in our analyses.

The analyses presented here were based on phenotypic records for individuals born between 1985 and 2012. Lambs were only included if their mother was dead to ensure that we were estimating lifetime space use for all animals. To prevent maternal rejections, we often delay weighing lambs until a few days after birth. As a result of early growth, the weight measurement will vary given the age at which they are caught. Because of this we restricted our birth weight analyses to individuals caught within five days of birth, and included capture age (in days) as a fixed effect in all birth weight models. We measured birth date as the number of days from 1 January, and August weight as the weight in kilograms of a lamb when it was caught in August. Jaw and metacarpal length measurements (in millimetres) were taken from bones that were collected and cleaned following mortality checks (see Beraldi et al. 2007 for more details), and in our analysis we only consider measurements taken from adult females (26 months or older) as skeletal growth is complete at this point (as indicated by an asymptote in the relationship between age and both jaw and metacarpal length [CER, unpublished results]).

Space use

We opted for two methods of accounting for space sharing within the animal model framework used to estimate the genetic parameters, which are broadly comparable to those used in Stopher et al. (2012) (differences are described below). The first involves directly accounting for SAC in the response variable, whilst the second involves quantifying home range similarity for pairs of individuals and incorporating this as an additional matrix. We started by extracting spatial information for each individual. We census the 170 hectare Village Bay area 30 times per year, 10 times in each of the three routine trips to the island (April-May, July-August, October-November). During each census, three fixed routes are walked simultaneously, the identity of all individuals seen is noted and their grid reference is recorded to the nearest 100 metres. We extracted lifetime census observations for all females, excluding any individuals that had fewer than 49 census observations in total. 49 observations is the minimum number needed to give an asymptote in core home range area, thereby providing a reliable estimate of the home range (see Regan et al. 2015 for details). We transformed these observations onto a grid, so that the most south-westerly census observation (NF091980) became (0,0) and each step on the grid represented a distance of 100 metres.

We next estimated individual space use. In the case of the SAC model each individual had to be assigned a single spatial location. We therefore calculated average lifetime locations for each female, ensuring that this was estimated to the nearest 100 metres corresponding to the grid described above. From these grid references we can consider SAC in either the East-West (column) or South-North axis (row), or both simultaneously.

To construct the home range similarity matrix necessary for the second method of accounting for space sharing in our animal models, we first estimated home ranges for each female. We estimated home ranges (100% isopleth) using kernel density methods, calculating the smoothing parameter using the *ad hoc* method, within the package adehabitatHR (Calenge, 2006). Because animals were assigned a grid reference to the nearest 100 metres during censuses, individuals frequently have numerous observations with identical grid references, and this can cause problems when estimating home ranges using kernel methods (Tufto et al., 1996). We therefore added a random number between -20 and 20 (representing a distance of up to 20 metres) to the X and Y coordinates for each record before estimating home ranges (see Moyes 2007 and Stopher et al. 2012). Powell (2000) suggests using core home ranges as they correspond to the area an animal uses most intensively, but here we were unable to do this because we could not construct a grid for home range estimation that was of a high enough resolution to give similarity metrics that scaled properly (i.e. between zero and one). We continued to consider only individuals with 49 or more observations, as doing so will still provide more reliable home range estimates, and similarity metrics. We then calculated home range overlap/similarity for all possible pairs of these females using Bhattacharyya's affinity (BA) (Bhattacharyya, 1943; Fieberg & Kochanny, 2005) in adehabitatHR (Calenge, 2006). We used BA (see Fieberg & Kochanny 2005 for a summary of possible metrics) for two reasons. Firstly, because it uses three dimensional utilisation distributions (UDs), which describe both where a home range is located in space and the probability of re-sighting an animal at different points within this home range, it better captures how individuals use different parts of their home range (Fieberg & Kochanny, 2005). Thus, this method provides more informative measures of similarity than metrics that consider only the spatial domain of the home ranges (Fieberg & Kochanny, 2005). Secondly it scales from zero (no overlap) to one (identical UDs), making it comparable to genetic relatedness, which is important when trying to tease apart the contributions of these sources of similarity. This provided us with a matrix containing pairwise similarity metrics for 931 females that could be incorporated into our models (see Fig. 1 for the distribution of BA values). In contrast to Stopher et al. (2012) we excluded individuals with insufficient census data in order to avoid potentially over- or under-estimating the bias caused by not accounting for space sharing.

Genomic relatedness matrix

When lambs are caught at birth they are sampled for genotyping. Individuals that are not caught at birth are sampled in August catches, by chemical immobilisation (darting, primarily of males during the rut), or post mortem. Genotypes at 37,037 informative autosomal single nucleotide polymorphism (SNP) markers on the Ovine SNP50 BeadChip (Illumina, for more information see Bérénos et al. 2014) are available for

5805 sheep spanning the period 1985-2012. The genomic relatedness between all pairs of SNP genotyped individuals was estimated in GCTA v1.04, which estimates the proportion of the genome identity-by-state between individuals (see Bérénos et al. 2014 for more details). This genomic relatedness matrix (GRM) was used in our animal models in place of the more traditional pedigree-derived additive relatedness matrix as it provides more accurate estimates of relatedness, leading to improved separation of direct and maternal genetic effects, and more precise estimates of quantitative genetic parameters (Bérénos et al., 2014).

Analyses

All analyses were conducted in R version 3.1.3 (R Development Core Team, 2008). We partitioned the phenotypic variance in each of the traits into genetic and environmental variance components using univariate animal models in ASReml-R (Butler et al., 2007). We included fixed effects to account for variation due to predictable effects such as sex and age. All models for early life traits included sex (two level factor), litter size (two level factor) and maternal age (linear and quadratic terms) as fixed effects. In addition, age at capture in days was included in models of birth weight (as a factor), and lamb August weight (as a covariate). For post mortem measures we only included a fixed effect of the age at death in months. After restricting on census observation number and removing individuals lacking the information needed to fit these fixed effects, we had birth weights for 1772 lambs (from 380 females), birth dates for 2124 lambs (404 females), August weights for 1043 lambs (334 females), and 300 and 286 females for jaw and metacarpal length analyses respectively.

We then added random effects sequentially. Firstly, we included a random effect of individual identity linked with the GRM to estimate the additive genetic effect (V_A, or the additive influence of genes carried by the individual in which the trait was measured). Secondly, we included a random effect of the year of birth, to partition the variance attributable to variation in the environment in the first year of life (V_{YoB}), followed by the identity of the individual's mother in order to estimate maternal effects (assuming that they are entirely environmental) (V_{ME}). Thirdly, in the case of the early life traits, we also fitted a maternal genetic effect (V_{MG}) to decompose the maternal effect variance into maternal permanent environment and maternal genetic components. This is important, as in the case of the early life traits we expect any bias caused by not accounting for space sharing by related females to be found in this maternal genetic effect component. Finally, we estimated the direct-maternal genetic covariance (COV_{am}) to enable the calculation of the total heritability.

We then accounted for space sharing in the following ways. Firstly, to account for spatial dependence in the response variable, we incorporated average lifetime locations by fitting column and row as additional

random effects, with an isotropic exponential covariance structure, equivalent to a continuous AR1 times AR1 process (Gilmour et al., 2009). This allows us to account for spatial autocorrelation between the residuals by dividing the residual error variance into spatially dependent and spatially independent residuals. It makes it possible to use an incomplete spatial array (where some intersections are not occupied by an individual) by including Column and row as random effects (Dutkowski et al., 2002), whilst also facilitating the examination of autocorrelation in Row and Column directions separately (Dutkowski et al., 2002). In models with the SAC process, we estimated both the variance explained by column and row ($V_{\text{Column and row}}$), and the strength of the autocorrelation (r). Secondly, we included information on home range similarity by fitting either individual identity (jaw and metacarpal length) or maternal identity (birth weight, birth date and lamb August weight) as an additional random effect, but this time linking it with our spatial similarity matrix (referred to subsequently as the '**S matrix**', with the corresponding variance component referred to as V_{Smatrix}).

The total phenotypic variance (denoted as Sum V in Table 2) was estimated as the sum of all variance components, and the variance explained by each of the variance components was calculated as the ratio of the relevant component to the total phenotypic variance. The direct additive-maternal genetic correlation (r_{am}) was calculated as $COV_{am}/\sqrt{V_A.V_{MG}}$. To account for maternal genetic effects and the direct additive-maternal genetic covariance when estimating heritability, we calculated the total heritability (h^2_T) as (V_A + $1.5COV_{am}$ + $0.5V_{MG}$)/Sum V (sensu Willham (1972), and following (Wilson et al., 2005)). We used likelihood ratio tests to assess the significance of random effects, assuming a χ^2 distribution with degrees of freedom equal to the number of additional parameters. However, because variance components cannot be smaller than zero (meaning the boundary condition is violated) the use of one degree of freedom can be overly conservative (Visscher, 2006). To gauge model credibility, we summed the variance component estimates from each model, with large changes in this total variance indicating potential problems with model performance, and that variance component estimates should be interpreted with some caution. In the Results, attention is drawn to models where this was the case, with the interpretation adjusted accordingly. For example code please see Appendix S1.

Results

Early life traits

We found evidence for strong maternal effects on all three early life traits, and models including maternal genetic effects (alongside maternal permanent environment effects and no spatial structure) performed signif-

icantly better than those estimating purely environmental maternal effects with no spatial structure (Birth weight - $\chi^2_{(df=1)}$=21.05, P< 0.001; Birth date - $\chi^2_{(df=1)}$=22.82, P< 0.001; August weight - $\chi^2_{(df=1)}$=14.12, P< 0.001). In fact, the estimate of the maternal genetic effect variance was consistently greater than that of the direct heritability for all three early life traits (Table 2). We did not however find any evidence for a significant direct-maternal genetic covariance for any of the three early life traits (Birth weight - $\chi^2_{(df=1)}$=0.001, P=0.97; Birth date - $\chi^2_{(df=1)}$=0.073, P=0.79; August weight - $\chi^2_{(df=1)}$=1.90, P=0.17).

We did however find some differences between these traits in the proportion of variance explained by the spatial term. For birth weight, inclusion of the **S matrix** significantly improved model fit ($\chi^2_{(df=1)}$=13.32, P< 0.001), and the term explained 5.6% of the variance (Table 2). Its incorporation resulted in small reductions in the estimates of V_{MG} (1.5% (from 16.9% to 15.4%), see Table 2) and V_{ME} (1.1% (from 2.9% to 1.8%), see Table 2), and therefore a negligible reduction in h^2_T (1.4% (from 9.4% to 8.0%), see Table 2). We found a similar trend when using the SAC models, again with a significant improvement in model fit when the spatial terms were added ($\chi^2_{(df=3)}$=10.56, P=0.014). The autocorrelation parameter indicated positive SAC (r=0.80), but column and row random effects only accounted for 3.6% of the variance and were associated with only a 2.3% reduction (from 16.9 to 14.6%) in the estimate of V_{MG}. The large standard errors, particularly around the estimate of the spatial variance component (Table 2), indicate that the model had some difficulty in estimating them, and lends credence to the idea that spatial variation in the environment does not generate substantial variation in lamb birth weight.

Including the **S matrix** also significantly improved model fit in the case of birth date ($\chi^2_{(df=1)}$=9.38, P=0.002). The spatial term accounted for 6.0% of the variance in birth date, and the change in the estimate of V_{MG} induced by not accounting for space sharing was higher than for birth weight, though still small, declining by 6.6% (from 25.5% to 18.9%) (Table 2, Fig. 2). The reduction in the maternal genetic effect estimate translated into a 3.4% decrease (from 17.1% to 13.7%) in the estimate of h^2_T (Table 2). When it came to the SAC model for birth date, we found evidence for a marginally significant improvement in model fit upon including column and row effects $\chi^2_{(df=3)}$=7.94, P=0.047), but there was substantial variance inflation upon incorporation of SAC, with the total variance explained increasing from 54.01 to 40935.84 (raw observed variance=48.971). Small changes in the variance explained are not of particular concern, but large changes may indicate that the model has produced poor parameter estimates (Stopher et al., 2012). This was associated with the spatial variance component becoming singular (Table 2), which prevented convergence. This suggests that there is too little spatial variance in the data to enable the estimation of the spatial parameters.

We also found that the **S matrix** significantly improved model fit when added to the model of August weight ($\chi^2_{(df=1)}$=10.12, P=0.001), but compared with the previous two traits, the spatial term accounted for

a much larger proportion of the total variance (17.8%, see Table 2 and Fig. 2). Despite this, the change in h^2_T caused by not accounting for the space sharing of females was on par with birth weight and birth date, with the estimate of V_{MG} reduced by only 4.9% (from 13.5% to 8.6%), and h^2_T reduced by 3.9% (Table 2). The results from the SAC model for August weight were similar to those from the birth date models. The model estimated a very large autocorrelation coefficient of 0.999 (Table 2), indicating very strong positive SAC in lamb August weight. However, the model failed to estimate the spatial variance component, with the estimate for this component increasing in size with each iteration before going singular (Table 2), and therefore preventing convergence. As in the case of birth date, this pattern may indicate that there is too little spatial variance in the data to enable the autocorrelation parameter to be estimated.

Adult traits

As expected from previous analyses, our estimates of h^2 (the ratio of V_A to the total trait variance) were much larger for jaw length and metacarpal length, than for the three early life traits (see Table 2), with only small amounts of variance attributable to birth year and maternal effects (see Table 2 and Fig. 2). We found a marginally significant improvement in model fit when adding the **S matrix** in the case of jaw length ($\chi^2_{(df=1)}$=3.96, P=0.046), with the term accounting for 8.2% of the variance in the trait (Table 2 and Fig. 2). The incorporation of the **S matrix** did result in a reduction in the estimate of h^2, though this was still relatively small at 6.8% (from 54.9% to 48.0%) (Table 2). For jaw length, the incorporation of SAC did not significantly improve model fit ($\chi^2_{(df=2)}$=1.87, P=0.599), explaining only 3.0% of the variance and resulting in only a 2.8% decrease (from 54.9% to 52.1%) in the estimate of h^2 (Table 2). In contrast to jaw length, adding the **S matrix** to models of metacarpal length did not improve model fit ($\chi^2_{(df=1)}$=0.11, P=0.74), and the term accounted for only 1.1% of the variance (Table 2 and Fig. 2). As a result we saw only a 1% reduction (from 77.1% to 76.1%) in the estimate of h^2 (Table 2). The SAC models echoed this result, as we saw no improvement in model fit upon the inclusion of SAC ($\chi^2_{(df=1)}$=-21.81, P>0.99), with it accounting for only 4.2% of the variance in the trait (Table 2 and Fig. 2). In addition, the very small autocorrelation parameter (that eventually went singular, see Table 2), suggests there is little evidence that animals which are similar in their space use are more similar in their metacarpal lengths than animals which range in very different parts of the study area. Please see Table 3 for fixed effects coefficients for each trait.

Discussion

As predicted, we found that increased similarity in female space use was associated with greater phenotypic similarity. This was most evident for the early life traits, with females that had more similar home ranges

having lambs that were more similar in their birth weights, birth dates, and August weights. Despite this, and contrary to our expectation, the increase in the (total) heritability estimates caused by not accounting for home range similarity was small, ranging from 1.4% (from 8.0% to 9.4%) to 6.8% (from 48.0% to 54.9%) depending on the trait considered.

Home range similarity generally explained a significant amount of variation in the traits considered, which aligns with previous research on the St. Kilda Soay sheep. Environmental components such as forage availability and quality vary markedly across the study area (Coulson et al., 1999; Regan et al., 2015). Such spatial variation in grazing quality means that sheep inhabiting different regions of the study area have access to food resources of differing quality, something that has been posited to lead to the variation in survival, recruitment and dispersal that we see across hefts (a heft being a group of individuals, regardless of sex or age, that use the same resources in space) (Coulson et al., 1999). Recent work has supported this idea, showing that variation in home range quality (measured as the home range percentage cover of *Holcus lanatus*, a key component of the *Holcus-Agrostis* plant community known to be highly palatable to the sheep (Crawley et al., 2004), is associated with variation in both male and female lifetime reproductive success (Regan et al., 2015). Given the heterogeneity in grazing quality across the study area, and the fact that females exhibit high fidelity to their natal heft (Coltman et al., 2003), it would follow that neighbouring animals are more phenotypically similar, particularly in traits such as birth weight and August weight. This is because these traits are likely to be strongly determined by the quantity and quality of food resources available to the mother during gestation and lactation (Oftedal, 1984). Though the **S matrix** improved model fit for all traits other than metacarpal length the proportion of variance explained by the spatial term was generally smaller than expected, particularly in the case of birth weight. There is one likely explanation for this result. Hirta's Soay sheep do not conform to the ideal free distribution (Jones 2006). Not only is *Holcus-Agrostis* grassland used by a greater proportion of the population than would be predicted from its availability, but its occupancy actually increases with sheep density (Jones 2006). This likely means that changes in population density compensate to some degree for the variation in grazing quality.

There are two conditions that need to be met for heritability estimates to be biased by disregarding the space use of animals. Firstly, relatives must be clustered in space as it is under this condition that phenotypic similarity due to shared genes may be confounded with similarity due to space sharing. The reason for this potential bias becomes clear when we consider variance partitioning methods. Assuming genetic and environment effects combine additively to determine phenotype, such that:

$$V_{\mathrm{P}} = V_{\mathrm{G}} + V_{\mathrm{E}}$$

It is apparent that such a model is only valid when there is no genotype-environment covariance. To meet this assumption any sources of correlation (i.e. common environment effects) must be accounted for elsewhere in the model. Social structure and natal philopatry are common in wild vertebrate populations, having been found in birds (Greenwood, 1980), mammals (Greenwood, 1980), reptiles (Sheridan et al., 2010), amphibians (Helfer et al., 2012), and fish (Mourier & Planes, 2013). As a result, the condition that relatives be clustered in space is likely to be satisfied for many natural populations. The degree of bias induced by failing to account for space sharing by relatives may however vary given the degree to which relatives are clustered in space, and the time scale over which the clustering is maintained. Complications may arise when considering migratory species, given that trait variation may be associated with conditions at either the wintering or breeding ground, or even both (Norris et al., 2004; Ockendon et al., 2013).

The second condition required for bias to occur is that the environment must be spatially heterogeneous, as it is this heterogeneity that will mean relatives are more likely to experience similar fine-scale environmental variation, and therefore appear more similar, than non-relatives. Again, this condition is likely to be satisfied in studies of natural populations, but the spatial scale, and pattern of this environmental heterogeneity is likely to be important, because it will influence the degree of environmental similarity experienced by relatives, compared to non-relatives. Though not a necessary condition for bias, trait choice should be carefully considered when drawing conclusions about the effect of including space sharing on heritability estimates. Accurately estimating heritability in quantitative genetic studies will necessitate the accounting of potential sources of common environment effects, such as space sharing by relatives. We therefore advocate for space sharing to be included, where the above conditions are met and where possible, into quantitative genetic analyses. However, it may be fruitful to focus on traits that, based on previous research, are believed to be heritable. Given the relatively limited knowledge surrounding the extent of the bias caused by space sharing it may be most economical to focus on the wide variety of traits for which substantial heritability is apparent in the literature. Furthermore, the degree of bias in heritability estimates as a result of failing to account for space sharing by related animals will be closely related to the degree of heritability in the trait. Despite fulfilling the above conditions and using traits believed to be heritable, we found no evidence of substantial bias in heritability estimates for any of the five traits studied. This suggests that these conditions alone are necessary but not sufficient to generate substantial bias in heritability estimates. Improving our understanding of philopatry in the St. Kilda Soay sheep will enable us to better put our findings into perspective. For example, we do not currently know how the associations between related individuals change over time. It is likely that these associations are not static, given that we know that female ranging behaviour changes with age in Soay sheep (Hayward et al., 2015). Similarly, it may be that dispersal varies across the study area, or between years, because of variation in habitat quality, resource availability or population density. Indeed,

dispersal is known to vary with environmental conditions in a wide range of species (Matthysen, 2005).

Our results suggest that although spatial effects can cause upward bias of heritability estimates, this is not always the case. This conclusion contrasts to those drawn in the two previous studies that have examined the change in heritability estimates when accounting for space sharing. In both cases they suggested that the bias was likely to be considerable, potentially up to 25% and 43%, respectively (Van Der Jeugd & McCleery, 2002; Stopher et al., 2012). Our estimates are likely to be more robust for the following reasons. Firstly, by using the animal model rather than parent-offspring regression (as used in Van Der Jeugd & McCleery (2002)), and a genomic relatedness matrix (GRM) instead of a traditional pedigree, the genetic parameter estimates are likely to be more precise (Kruuk, 2004; Akesson et al., 2008; Bérénos et al., 2014)(though note that animal models were used in Stopher et al. 2012). Secondly, when it came to the **S matrix** approach, we only estimated home ranges, and calculated similarity metrics, for females with at least 49 census observations. Kernel density methods are sensitive to the availability of location data (Seaman et al., 1999; Blundell et al., 2001), and we wanted to ensure that our spatial estimates were not influenced by poor home range estimates for individuals with few data, and assumptions of zero overlap for individuals with no data. It is likely that in other studies spatial data will be more limiting than in our case. The number of observations necessary to accurately estimate home ranges will however vary between systems and with the method in which the spatial data were collected. It will therefore be important to calculate the likely number of observations needed for accurate home range estimation on a case-by-case basis. Where smaller spatial datasets are available it may be possible to run the analyses with subsets of individuals that vary in their number of observations in order to understand how this influences results. In addition, it may be possible to use tools such as Bayesian kernel density estimators (Zhang et al., 2006) to appropriately account for the uncertainty surrounding home range estimates when deriving overlap metrics. Fourth, and finally, by choosing traits that based on previous research were believed to be heritable we hope our results will provide a useful benchmark for further studies of this kind. As mentioned above, trait choice is likely to be important when drawing conclusions about the expected change in heritability estimates upon accounting for the space sharing of relatives. For example, the large change in the estimated heritability for home range size traits upon including space sharing that was uncovered by Stopher et al. (2012) suggested that the bias in heritability estimates is likely to be substantial. These traits are very likely to be spatially variable given their close link with food availability (Tufto et al., 1996; Eide et al., 2004). This made them ideal for illustrating the bias that could be expected under a worse-case scenario but such traits are unlikely to yield representative estimates of the degree of bias in quantitative genetic parameters because there is, to our knowledge, little evidence to suggest that they have a heritable basis, particularly in mammals. This suggestion is supported by the fact that the results from **S matrix** models of birth weight were largely comparable between our study and that of Stopher et al. (2012). Both

the proportion of variance explained by the spatial term (5.6% in our analysis, and 5.9% in Stopher et al. (2012)), and the change in the estimate of h^2 when not accounting for shared space (1.4% (from 9.4% to 8.0%) in our analysis, and 2.6% (from 28.2% to 25.6%) in Stopher et al. (2012)) were of similar magnitude.

The results from SAC models for birth weight reported here were less similar to those of Stopher et al. (2012). In our analysis, the results from the two methods were generally comparable, but in the analysis by Stopher et al. (2012), incorporation of the SAC process resulted in the proportion of variance explained by the spatial term increasing to 19.5% (from 5.9% when using the **S matrix**), and an absolute change in the estimate of h^2 of 14.4%. As Stopher et al. (2012) suggest, their results may indicate that different spatial processes are at work, but there was some indication that their SAC model could not estimate the autocorrelation coefficient, given that it was fixed at the boundary. Furthermore, the standard error around the variance component estimate for the spatial term was very large (estimate=0.336, standard error=0.700), suggesting that there may be little spatial variation in birth weight. Consequently, our results appear to be more closely aligned with those of Stopher et al. (2012) than it may at first seem.

In light of this work we make some recommendations for future studies aiming to account for space sharing by relatives when running quantitative genetic analyses. In some cases our SAC models poorly estimated the autocorrelation parameter and the variance explained by SAC. Though these models can indicate whether there is spatial dependence in a trait, it is difficult to put weight on the estimates of the spatial variance component, and therefore on the change in the estimated heritability. The problems we and Stopher et al. (2012) have identified with the SAC models is perhaps unsurprising, given that they were developed for the analysis of agricultural variety trials (Cullis & Gleeson, 1991). The data from such trials differ considerably to those from long-term studies of animal populations, and this has the potential to influence the suitability of SAC approaches. For example, crop and forestry trials deal with non-mobile organisms, that have single spatial locations. We can assign animals single locations, making it possible to run SAC models, but this may reduce our ability to detect a spatial signal, having averaged over detailed information on individual space use. Furthermore, by averaging over each individual's location data multiple animals often have the same average location (at least over the spatial scale we were able to work at), despite the fact that they do vary in their space use. This too could make it difficult to detect a spatial trend. Finally, we can often record the locations of plants at a much higher precision than that of wild animals. For example, in the case of the Soay sheep, census data are only recorded to the nearest 100 metres. Therefore, when it comes to studies of wild animals, there is often a much courser grid over which to run these analyses. This is likely to be one of the key reasons for the poor estimation of the autocorrelation parameter and/or the spatial variance component in our SAC models. As a result, we agree with Stopher et al. (2012) in advocating the **S matrix** approach. Not only is it relatively straightforward to fit, but it is arguably the best available way of

including information on space use similarity in animal models. This is because it uses similarity metrics that are based on three dimensional utilisation distributions, which tell us not only where a home range is located, but actually to what degree animals use different parts of this home range (Worton, 1989). One potential limitation of the **S matrix** approach is that, by capturing information on home range overlap/similarity, it can say nothing about individuals that live adjacent to each other but have non-overlapping utilisation distributions according to Bhattacharyya's Affinity. Though it is unlikely that two animals could live at close quarters without overlapping at all in their distributions, it may be inappropriate to leave SAC models behind altogether, because they provide a means to capture this information.

Though our work shows that quantitative geneticists may have confidence in their heritability estimates, there is some way to go before we can make informed predictions about the degree of bias in heritability estimates when we cannot, or do not account for spatially derived similarity. As a result, a key avenue for future research is in understanding whether the degree of bias varies between species, given the huge variation in dispersal patterns in nature. This will make it possible to predict the need for spatial components in quantitative genetic models in the future. It is also important to consider precisely what aspect of the environment is varying spatially when conducting these analyses. In our study, we were generally concerned with capturing the effect of variation in resource availability, with such variation likely to impact traits associated with growth. In other studies, however, the focus may be on spatial variation in predation risk due to variation in substrate colour or vegetation structure, or even spatial variation in the social environment due to differences in, for example, density. This focus will dictate which traits are of most interest, or where bias is of greatest concern.

There are two other exciting avenues for research that we wish to draw attention to. Firstly, though we accounted for phenotypic similarity caused by individuals being born in the same year, the current models lack information regarding temporal variation in the environment after the year of birth and temporal changes in space use itself. Currently, these models treat individuals with a given home range overlap, or neighbouring average lifetime locations, equivalently, whether they were alive at the same time or their lives never overlapped. This assumption of equality regardless of temporal overlap is probably over-simplifying, and penalising the similarity metric for individuals whose lives did not overlap might result in smaller changes in heritability estimates upon including space sharing. Ranging behaviour itself may vary temporally, and therefore it may be necessary to consider the temporal scale at which space sharing is quantified more carefully. For example, early life traits may be more dependent on environmental variation at a temporal scale below that of the lifetime, because the body mass/condition of adult females is likely determined at this scale. Indeed, a number of mammalian studies have shown variation in adult body mass/condition in relation to temporal variation in the environment (Clutton-Brock & Albon, 1983; Toïgo et al., 2006). It

may therefore be preferable to analyse such traits using an **S matrix** constructed at a more appropriate temporal scale. The suitability of this approach will, however, depend on data availability, as animal models are necessarily data hungry. Secondly, though perhaps unlikely in mammals, traits associated with ranging behaviour may have a genetic basis. Indeed there have now been a number of quantitative genetic studies focusing on traits associated with ranging behaviour and dispersal (Waser & Thomas Jones, 1989; Hansson et al., 2003; Doligez et al., 2009; Charmantier et al., 2011), the majority of which have focused on birds. If there is a genetic element to space use itself, then it is possible that by accounting for space sharing of related individuals heritabilities may be underestimated (Stopher et al., 2012).

In conclusion, we find that despite significant spatial variation in a variety of heritable traits there were only small changes in heritability estimates when we failed to account for the fact that related female Soay sheep share space because of natal philopatry. This suggests that heritability estimates from prior quantitative genetic studies of this population are likely to be reliable. Though this is reassuring further research will be needed before we can be confident of the generality of these results. We hope that this work will encourage researchers to include spatial processes in their animal models when their data fulfil the conditions under which we would expect bias in heritability estimates by not accounting for space sharing. Not only that, we hope that they will publish their results, even when heritability estimates are largely unchanged, so that we can better predict when bias may be of particular concern.

References

Akesson, M., Bensch, S., Hasselquist, D., Tarka, M. & Hansson, B. 2008. Estimating heritabilities and genetic correlations: comparing the 'animal model' with parent-offspring regression using data from a natural population. *PLoS One* **3**: 1–10.

Banerjee, S., Finley, A.O., Waldmann, P. & Ericsson, T. 2010. Hierarchical Spatial Process Models for Multiple Traits in Large Genetic Trials. *J. Am. Stat. Assoc.* **105**: 506–521.

Beraldi, D., McRae, A.F., Gratten, J., Slate, J., Visscher, P.M. & Pemberton, J.M. 2007. Mapping quantitative trait loci underlying fitness-related traits in a free-living sheep population. *Evolution* **61**: 1403–16.

Bérénos, C., Ellis, P.A., Pilkington, J.G. & Pemberton, J.M. 2014. Estimating quantitative genetic parameters in wild populations: A comparison of pedigree and genomic approaches. *Mol. Ecol.* **23**: 3434–3451.

Bhattacharyya, A. 1943. On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutte. Math. Soc* **35**: 99–109.

Blundell, G.M., Maier, J.A.K. & Debevec, E.M. 2001. Linear home ranges: Effects of smoothing, sample size, and autocorrelation on kernel estimates. *Ecol. Monogr.* **71**: 469–489.

Butler, D.G., Cullis, B.R., Gilmour, A.R. & Gogel, B.J. 2007. Analysis of Mixed Models for S Language Environments: ASReml-R Reference Manual, Release 2. Queensland Department of Primary Industries and Fisheries, Brisbane, Australia.

Calenge, C. 2006. The package adehabitat for the R software: A tool for the analysis of space and habitat use by animals. *Ecol. Model.* **197**: 516–519.

Charmantier, A., Buoro, M., Gimenez, O. & Weimerskirch, H. 2011. Heritability of short-scale natal dispersal in a large-scale foraging bird, the wandering albatross. *J. Evol. Biol.* **24**: 1487–1496.

Cliff, A. & Ord, J. 1981. *Spatial processes: models and applications.* Pion, London.

Clutton-Brock, T.H. & Albon, S.D. 1983. Climatic Variation and Body Weight of Red Deer. *J. Wildl. Manage.* **47**: 1197–1201.

Clutton-Brock, T.H., Pemberton, J.M., Coulson, T., Stevenson, I.R. & Maccoll, A.D.C. 2004. Individuals and populations. In: *Soay sheep: dynamics and selection in an island population* (T.H. Clutton-Brock & J.M. Pemberton, eds), pp. 1–13. Cambridge University Press, Cambridge, United Kingdom.

Coltman, D.W., Pilkington, J.G. & Pemberton, J.M. 2003. Fine-scale genetic structure in a free-living ungulate population. *Mol. Ecol.* **12**: 733–742.

Coulson, T., Albon, S., Pilkington, J. & Clutton-Brock, T. 1999. Small-scale spatial dynamics in a fluctuating ungulate population. *J. Anim. Ecol.* **68**: 658–671.

Crawley, M.J., Albon, S.D., Bazely, D.R., Milner, J.M., Pilkington, J.G. & Tuke, A.L. 2004. Vegetation and sheep population dynamics. In: *Soay sheep: dynamics and selection in an island population* (T.H. Clutton-Brock & J.M. Pemberton, eds), pp. 89–111. Cambridge University Press, Cambridge, United Kingdom.

Cullis, B.R. & Gleeson, A.C. 1989. Efficiency of neighbour analysis for replicated variety trials in Australia. *J. Agric. Sci.* **113**: 233–239.

Cullis, B.R. & Gleeson, A.C. 1991. Spatial Analysis of Field Experiments-An Extension to Two Dimensions. *Biometrics* **47**: 1449–1460.

Doligez, B., Gustafsson, L. & Pärt, T. 2009. 'Heritability' of dispersal propensity in a patchy population. *Proc. R. Soc. Lond. B. Biol. Sci.* **276**: 2829–2836.

Dutkowski, G.W., Costa e Silva, J., Gilmour, A.R., Wellendorf, H. & Aguiar, A. 2006. Spatial analysis enhances modelling of a wide variety of traits in forest genetic trials. *Can. J. For. Res.* **36**: 1851–1870.

Dutkowski, G.W., Silva, J.C.e., Gilmour, A.R. & Lopez, G.A. 2002. Spatial analysis methods for forest genetic trials. *Can. J. For. Res.* **32**: 2201–2214.

Eide, N., Jepsen, J. & Prestrud, P. 2004. Spatial organization of reproductive Arctic foxes Alopex lagopus: responses to changes in spatial and temporal availability of prey. *J. Anim. Ecol.* **73**: 1056–1068.

Falconer, D. & Mackay, T.F. 1996. *Introduction to quantitative genetics. 4th Edition.* Prentice Hall, London.

Fieberg, J. & Kochanny, C. 2005. Quantifying home-range overlap: the importance of the utilization distribution. *J. Wildl. Manage.* **69**: 1346–1359.

Gilmour, A.R., Gogel, B., Cullis, B.R. & Thompson, R. 2009. Asreml User Guide Release 3.0.

Greenwood, P.J. 1980. Mating systems, philopatry and dispersal in birds and mammals. *Anim. Behav.* **28**: 1140–1162.

Hadfield, J.D., Burgess, M.D., Lord, A., Phillimore, A.B., Clegg, S.M. & Owens, I.P.F. 2006. Direct versus indirect sexual selection: genetic basis of colour, size and recruitment in a wild bird. *Proc. R. Soc. Lond. B. Biol. Sci.* **273**: 1347–53.

Hansson, B., Bensch, S. & Hasselquist, D. 2003. Heritability of dispersal in the great reed warbler. *Ecol. Lett.* **6**: 290–294.

Hayward, A.D., Moorad, J., Regan, C.E., Berenos, C., Pilkington, J.G., Pemberton, J.M. & Nussey, D.H. 2015. Asynchrony of senescence among phenotypic traits in a wild mammal population. *Exp. Gerontol.* **71**: 56–68.

Heckerman, D., Gurdasani, D., Kadie, C., Pomilla, C., Carstensen, T., Martin, H., Ekoru, K., Nsubuga, R.N., Ssenyomo, G., Kamali, A., Kaleebu, P., Widmer, C. & Sandhu, M.S. 2016. Linear mixed model for heritability estimation that explicitly addresses environmental variation. *Proc. Natl. Acad. Sci.* **113**: 7377–7382.

Helfer, V., Broquet, T. & Fumagalli, L. 2012. Sex-specific estimates of dispersal show female philopatry and male dispersal in a promiscuous amphibian, the alpine salamander (Salamandra atra). *Mol. Ecol.* **21**: 4706–4720.

Kie, J.G., Matthiopoulos, J., Fieberg, J., Powell, R.A., Cagnacci, F., Mitchell, M.S., Gaillard, J.M. & Moorcroft, P.R. 2010. The home-range concept: are traditional estimators still relevant with modern telemetry technology? *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **365**: 2221–31.

Kruuk, L.E., Slate, J. & Wilson, A.J. 2008. New Answers for Old Questions: The Evolutionary Quantitative Genetics of Wild Animal Populations. *Annu. Rev. Ecol. Evol. Syst.* **39**: 525–548.

Kruuk, L.E.B. 2004. Estimating genetic parameters in natural populations using the "animal model". *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **359**: 873–890.

Kruuk, L.E.B. & Hadfield, J.D. 2007. How to separate genetic and environmental causes of similarity between relatives. *J. Evol. Biol.* **20**: 1890–1903.

Lynch, M. & Walsh, B. 1998. *Genetics and analysis of quantitative traits.* Sinauer, Sunderland.

Magnussen, S. 1993. Bias in genetic variance estimates due to spatial autocorrelation. *Theor. Appl. Genet.* **86**: 349–55.

Matthysen, E. 2005. Density-dependent dispersal in birds and mammals. *Ecography* **28**: 403–416.

McAdam, A.G., Boutin, S., Réale, D. & Berteaux, D. 2002. Maternal effects and the potential for evolution in a natural population of animals. *Evolution* **56**: 846–851.

McCleery, R., Pettifor, R., Armbruster, P., Meyer, K., Sheldon, B. & Perrins, C. 2004. Components of Variance Underlying Fitness in a Natural Population of the Great Tit Parus major. *Am. Nat.* **164**: E62–E72.

Mourier, J. & Planes, S. 2013. Direct genetic evidence for reproductive philopatry and associated fine-scale migrations in female blacktip reef sharks (Carcharhinus melanopterus) in French Polynesia. *Mol. Ecol.* **22**: 201–214.

Moyes, K. 2007. *Demographic consequences of individual variation.* Ph.d. thesis, University of Kent, Canterbury, U.K.

Ng, J., Landeen, E.L., Logsdon, R.M. & Glor, R.E. 2013. Correlation between Anolis lizard deqlap phenotype and environmentl variation indicates adaptive divergence of a signal important to sexual selection and species recognition. *Evolution* **67**: 573–582.

Nilsson, J. & Svensson, M. 1996. Sibling Competition Affects Nestling Growth Strategies in Marsh Tits. *J. Anim. Ecol.* **65**: 825–836.

Norris, D.R., Marra, P.P., Kyser, T.K., Sherry, T.W. & Ratcliffe, L.M. 2004. Tropical winter habitat limits reproductive success on the temperate breeding grounds in a migratory bird. *Proc. R. Soc. Lond. B. Biol. Sci.* **271**: 59–64.

Ockendon, N., Leech, D. & Pearce-Higgins, J.W. 2013. Climatic effects on breeding grounds are more important drivers of breeding phenology in migrant birds than carry-over effects from wintering grounds. *Biol. Lett.* **9**: 1–4.

Oftedal, O.T. 1984. Body size and reproductive strategy as correlates of milk energy output in lactating mammals. *Acta Zool. Fenn.* **171**: 183–186.

Olalla-Tárraga, M.Á., Rodríguez, M.Á. & Ángel, M. 2007. Energy and interspecific body size patterns of amphibian faunas in Europe and North America: anurans follow Bergmann's rule, urodeles its converse. *Glob. Ecol. Biogeogr.* **16**: 606–617.

Powell, R. 2000. Animal home ranges and territories and home range estimators. pp. 65–110. Columbia University Press, New York, USA.

Qiao, C.G., Basford, K.E., DeLacy, I.H. & Cooper, M. 2000. Evaluation of experimental designs and spatial analyses in wheat breeding trials. *Theor. Appl. Genet.* **100**: 9–16.

R Development Core Team 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Regan, C.E., Pilkington, J.G., Pemberton, J.M. & Crawley, M.J. 2015. Sex differences in relationships between habitat use and reproductive performance in Soay sheep (Ovis aries). *Ecol. Lett.* **19**: 171–179.

Rossiter, S.J., Jones, G., Ransome, R.D. & Barratt, E.M. 2002. Relatedness structure and kin-biased foraging in the greater horseshoe bat (Rhinolophus ferrumequinum). *Behav. Ecol. Sociobiol.* **51**: 510–518.

Seaman, D., Millspaugh, J., Kernohan, B., Brundige, G., Raedeke, K. & Gitzen, R. 1999. Effects of Sample Size on Kernel Home Range Estimates. *J. Wildl. Manag.* **63**: 739–747.

Sheridan, C.M., Spotila, J.R., Bien, W.F. & Avery, H.W. 2010. Sex-biased dispersal and natal philopatry in the diamondback terrapin, Malaclemys terrapin. *Mol. Ecol.* **19**: 5497–510.

Silva, J.C.e., Dutkowski, G.W. & Gilmour, A.R. 2001. Analysis of early tree height in forest genetic trials is enhanced by including a spatially correlated residual. *Can. J. For. Res.* **31**: 1887–1893.

Stopher, K.V., Walling, C.A., Morris, A., Guinness, F.E., Clutton-Brock, T.H., Pemberton, J.M. & Nussey, D.H. 2012. Shared spatial effects on quantitative genetic parameters: Accounting for spatial autocorrelation and home range overlap reduces estimates of heritability in wild red deer. *Evolution* **66**: 2411–2426.

Toïgo, C., Gaillard, J.M., Van Laere, G., Hewison, M. & Morellet, N. 2006. How does environmental variation influence body mass, body size, and body condition? Roe deer as a case study. *Ecography* **29**: 301–308.

Tufto, J., Andersen, R. & Linnell, J. 1996. Habitat use and ecological correlates of home range size in a small cervid : the roe deer. *J. Anim. Ecol.* **65**: 715–724.

van Beest, F.M., Rivrud, I.M., Loe, L.E., Milner, J.M. & Mysterud, A. 2011. What determines variation in home range size across spatiotemporal scales in a large browsing herbivore? *J. Anim. Ecol.* **80**: 771–785.

Van Der Jeugd, H.P. & McCleery, R. 2002. Effects of spatial autocorrelation, natal philopatry and phenotypic plasticity on the heritability of laying date. *J. Evol. Biol.* **15**: 380–387.

Vergara, P., Fargallo, J.A. & Martínez-Padilla, J. 2015. Genetic basis and fitness correlates of dynamic carotenoid-based ornamental coloration in male and female common kestrels Falco tinnunculus. *J. Evol. Biol.* **28**: 146–54.

Visscher, P.M. 2006. A note on the asymptotic distribution of likelihood ratio tests to test variance components. *Twin Res. Hum. Genet.* **9**: 490–5.

Waser, P.M. & Thomas Jones, W. 1989. Heritability of dispersal in banner-tailed kangaroo rats, Dipodomys spectabilis. *Anim. Behav.* **37**: 987–991.

Willham, R.L. 1972. The role of maternal effects in animal breeding. 3. Biometrical aspects of maternal effects in animals. *J. Anim. Sci.* **35**: 1288–93.

Wilson, A.J., Coltman, D.W., Pemberton, J.M., Overall, A.D.J., Byrne, K.A. & Kruuk, L.E.B. 2005. Maternal genetic effects set the potential for evolution in a free-living vertebrate population. *J. Evol. Biol.* **18**: 405–14.

Wilson, A.J., Pemberton, J.M., Pilkington, J.G., Clutton-Brock, T.H., Coltman, D.W. & Kruuk, L.E.B. 2007. Quantitative genetics of growth and cryptic evolution of body size in an island population. *Evol. Ecol.* **21**: 337–356.

Wilson, A.J., Réale, D., Clements, M.N., Morrissey, M.M., Postma, E., Walling, C.A., Kruuk, L.E.B. & Nussey, D.H. 2010. An ecologist's guide to the animal model. *J. Anim. Ecol.* **79**: 13–26.

Worton, B. 1989. Kernel methods for estimating the utilization distribution in home-range studies. *Ecology* **70**: 164–168.

Zhang, X., King, M.L. & Hyndman, R.J. 2006. A Bayesian approach to bandwidth selection for multivariate kernel density estimation. *Comput. Stat. Data Anal.* **50**: 3009–3031.

Table 1: Published estimates of the narrow-sense heritability (h^2), maternal genetic effect variance (m^2), and total heritability (h^2_T, when reported), for the five traits considered in this study. Standard errors are provided in parentheses, where available. Note though that the fixed effects included in our models were similar to (but not identical) those included in models in these analyses. Because variance component ratios were calculated using the sum of the variance components as the denominator, reported heritabilities are conditional on fixed effects.

Trait	h^2	m^2	h^2_T	Reference
Birth weight (lamb)	0.075 (0.045)	0.119 (0.045)	0.135(0.045)	Wilson et al. (2005)
	0.160	0.250		Beraldi et al. (2007)
	0.069	0.284	-	Wilson et al. (2007)
	0.059 (0.017)	0.155 (0.033)	-	Bérénos et al. (2014)
Birth date (lamb)	0.055 (0.036)	0.283 (0.051)	0.197 (0.038)	Wilson et al. (2005)
	0.070	0.690		Beraldi et al. (2007)
August weight (lamb)	0.047	0.017	-	Wilson et al. (2007)
	0.104 (0.026)	0.103 (0.032)	-	Bérénos et al. (2014)
Jaw length (adult female)	0.390	-	-	Beraldi et al. (2007)
	0.594 (0.070)	-	-	Bérénos et al. (2014)
Metacarpal length (adult female)	0.450	-	-	Beraldi et al. (2007)
	0.556 (0.072)	-	-	Bérénos et al. (2014)

Table 2: Variance component estimates and their associated ratios for models including no spatial term, with the **S matrix** (containing home range similarity metrics), or with spatial autocorrelation, for three early life, and two adult traits measured in St. Kilda Soay sheep. Reported are the additive genetic variance (V_A), birth year variance (V_{YoB}), maternal permanent environment variance (V_{ME}), maternal genetic variance (V_{MG}), **S matrix** variance ($V_{Smatrix}$), SAC variance ($V_{Column\ and\ row}$), the total variance (Sum V), the autocorrelation coefficient (Autocorrelation (r)), the direct additive-maternal genetic covariance (COV_{am}) and correlation (r_{am}), and the total heritability (h^2_T, accounting for maternal genetic effects and the direct additive-maternal genetic covariance). We provide both the raw variance component/correlation estimates ('Est'), and the proportion of the total variance explained by the term in the case of variance components ('Prop'). Standard errors are provided in parentheses, and singular parameters (with 'NA' standard errors)/parameters that were fixed at the boundary are italicised

	No spatial term		With S matrix		With SAC	
	Est	Prop	Est	Prop	Est	Prop
Birth weight						
V_A	0.003 (0.004)	0.014 (0.018)	0.003 (0.004)	0.014 (0.017)	0.003 (0.004)	0.011 (0.017)
V_{YoB}	0.081 (0.024)	0.336 (0.069)	0.081 (0.025)	0.324 (0.069)	0.081 (0.025)	0.331 (0.070)
V_{ME}	0.007 (0.007)	0.029 (0.030)	0.004 (0.007)	0.018 (0.027)	0.008 (0.007)	0.032 (0.028)
V_{MG}	0.041 (0.011)	0.169 (0.047)	0.038 (0.011)	0.154 (0.043)	0.033 (0.010)	0.135 (0.042)
$V_{Smatrix}$			0.014 (0.010)	0.056 (0.040)		
$V_{Column\ and\ row}$					0.010 (0.012)	0.039 (0.048)
$V_{Residual}$	0.108 (0.005)	0.451 (0.053)	0.108 (0.005)	0.436(0.052)	0.111 (0.005)	0.451 (0.056)
Sum V	0.240		0.249		0.246	
Autocorrelation (r)					0.798 (0.305)	
COV_{am}	$-8.830\times e^{4}$ (0.006)		-0.002 (0.005)		$-5.037\times e^{-4}$ (0.005)	
r_{am}	-0.075 (0.445)		-0.158 (0.427)		-0.053 (0.540)	
h^2_T	0.094 (0.030)		0.080 (0.027)		0.076 (0.028)	
Birth date						
V_A	1.740 (1.077)	0.032 (0.020)	1.759 (1.085)	0.032 (0.020)	1.722 (1.079)	$7.911\times e^{-5}$ ($7.740\times e^{-5}$)
V_{YoB}	7.245 (2.208)	0.134 (0.036)	7.282 (2.220)	0.133 (0.036)	7.376 (2.245)	$3.390\times e^{-4}$ ($2.728\times e^{-4}$)
V_{ME}	7.478 (2.608)	0.138 (0.049)	8.231 (2.559)	0.150 (0.048)	8.457 (2.589)	$3.886\times e^{-4}$ ($3.112\times e^{-4}$)
V_{MG}	13.757 (3.810)	0.255 (0.063)	10.366 (3.503)	0.189 (0.060)	10.757 (3.540)	$4.943\times e^{-4}$ ($4.144\times e^{-4}$)
$V_{Smatrix}$			3.314 (2.776)	0.060 (0.048)		
$V_{Column\ and\ row}$					21708.86 (16248.195)	0.998 (0.002)
$V_{Residual}$	23.793 (1.105)	0.441 (0.035)	23.912 (1.112)	0.436 (0.038)	23.914 (1.111)	0.001 ($8.216\times e^{-4}$)
Sum V	54.014		54.864		21761.08	
Autocorrelation (r)					*1.000(NA)*	
COV_{am}	0.455 (1.539)		0.417 (1.462)		0.176 (1.469)	
r_{am}	0.093 (0.327)		0.098 (0.357)		0.041 (0.347)	
h^2_T	0.171 (0.042)		0.137 (0.041)		$3.384\times e^{-4}$ ($2.802\times e^{-4}$)	

August weight						
V_A	0.198 (0.192)	0.036 (0.035)	0.170 (0.189)	0.029 (0.032)	0.163 (0.186)	$2.872\times e^{-5}$ ($3.431\times e^{-5}$)
V_{YoB}	2.129 (0.658)	0.391 (0.075)	1.670 (0.527)	0.287 (0.074)	1.970 (0.609)	$3.473\times e^{-4}$ ($1.111\times e^{-4}$)
V_{ME}	0.021 (0.191)	0.004 (0.035)	0.043 (0.176)	0.007 (0.030)	0.054 (0.180)	$9.449\times e^{-6}$ ($3.190\times e^{-5}$)
V_{MG}	0.737 (0.291)	0.135 (0.052)	0.498 (0.253)	0.086 (0.044)	0.532 (0.260)	$9.386\times e^{-5}$ ($4.693\times e^{-5}$)
$V_{Smatrix}$			1.032 (0.629)	0.178 (0.093)		
$V_{Column\ and\ row}$					*5665.361 (428.498)*	*0.999 ($1.327\times e^{-4}$)*
$V_{Residual}$	2.359 (0.182)	0.433 (0.065)	2.397 (0.183)	0.413 (0.064)	2.405 (0.182)	$4.241\times e^{-4}$ ($5.631\times e^{-8}$)
Sum V	5.445		5.810		5670.48	
Autocorrelation (r)					1.000 ($1.083\times e^{-5}$)	
COV_{am}	0.244 (0.169)		0.222 (0.158)		0.226 (0.157)	
r_{am}	0.638 (0.658)		0.764 (0.871)		0.769 (0.865)	
h^2_T	0.164 (0.046)		0.125 (0.041)		$1.355\times e^{-4}$ ($4.629\times e^{-5}$)	
Jaw length						
V_A	9.974 (2.834)	0.549 (0.129)	9.045 (2.726)	0.480 (0.129)	9.701 (2.799)	0.531 (0.131)
V_{YoB}	*$4.659\times e^{-7}$ ($2.332\times e^{-7}$)*	*$2.562\times e^{-8}$ ($1.376\times e^{-8}$)*	*$5.177\times e^{-7}$ ($2.324\times e^{-8}$)*	*$2.747\times e^{-8}$ ($1.336\times e^{-8}$)*	*$4.834\times e^{-7}$ ($2.344\times e^{-7}$)*	*$2.644\times e^{-8}$ ($1.380\times e^{-8}$)*
V_{ME}	3.603 (1.761)	0.198 (0.092)	3.137 (1.724)	0.166 (0.089)	3.223 (1.750)	0.176 (0.092)
$V_{Smatrix}$			1.548 (1.493)	0.082 (0.074)		
$V_{Column\ and\ row}$					0.584 (0.980)	0.031 (0.052)
$V_{Residual}$	4.604 (2.305)	0.253 (0.136)	5.116 (2.297)	0.271 (0.132)	4.777 (2.316)	0.261 (0.136)
Sum V	18.180		18.846		18.285	
Autocorrelation (r)					0.679 (0.773)	
Metacarpal length						
V_A	10.340 (2.741)	0.771 (0.110)	13.306 (2.744)	0.761 (0.114)	9.533 (1.509)	0.408 (0.066)
V_{YoB}	*$3.008\times e^{-7}$ ($3.138\times e^{-7}$)*	*$1.731\times e^{-8}$ ($1.878\times e^{-8}$)*	*$1.946\times e^{-7}$ ($1.987\times e^{-7}$)*	*$1.113\times e^{-8}$ ($1.182\times e^{-8}$)*	*0.004(0.451)*	*$1.758\times e^{-4}$ (0.019)*
V_{ME}	2.094 (1.402)	0.121 (0.079)	2.054 (1.397)	0.118 (0.078)	2.487 (1.553)	0.106 (0.068)
$V_{Smatrix}$			0.020 (0.068)	0.011 (0.039)		
$V_{Column\ and\ row}$					1.069 (0.861)	0.046 (0.036)
$V_{Residual}$	1.885 (1.966)	0.108 (0.118)	1.923 (1.963)	0.110 (0.117)	10.266 (2.472)	0.439 (0.069)
Sum V	17.379		17.479		23.359	
Autocorrelation (r)					*$2.148\times e^{-69}$(NA)*	

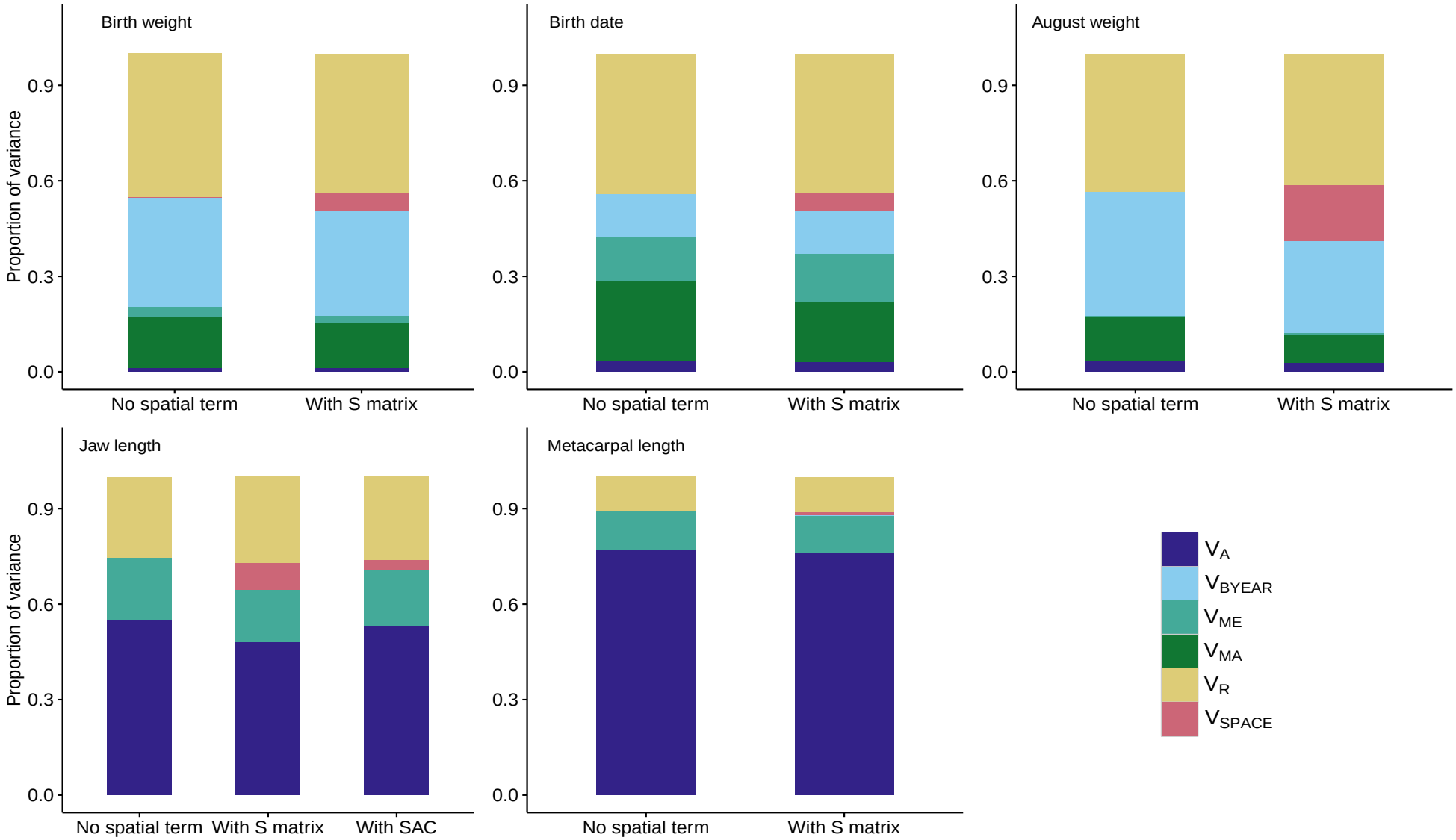

Figure 1: The proportions of variance explained by different random effects in animal models for three early life, and two adult traits measured in St. Kilda Soay sheep. For all traits we present the results for models containing no spatial term, and with the home range similarity matrix (or '**S matrix**'), however we only present the results from spatial autocorrelation models ('With SAC') for the traits where SAC models converged

Table 3: Coefficients and standard errors of fixed effects for each of the five traits featured in this study. For early life traits, values are taken from a model including only the fixed effects shown in the table, and additive genetic, birth year, maternal permanent environment, maternal genetic random effects and a direct-maternal genetic covariance. For the adult skeletal traits, values are taken from a model including only the fixed effects below, and additive genetic, birth year, and maternal permanent environment random effects.

Trait	Fixed effect	Level	Coefficient	Standard Error
Birth weight				
	Litter size	Singleton	0.000	NA
	Litter size	Twin	-0.795	0.023
	Sex	Female	0.000	NA
	Sex	Male	0.119	0.017
	Maternal age		0.438	0.014
	Maternal age^2		-0.031	0.001
	Capture age	Day zero	0.000	NA
	Capture age	Day one	0.245	0.042
	Capture age	Day two	0.397	0.041
	Capture age	Day three	0.481	0.044
	Capture age	Day four	0.518	0.049
	Capture age	Day five	0.552	0.057
Birth date				
	Litter size	Singleton	0.000	NA
	Litter size	Twin	0.003	0.330
	Sex	Female	0.000	NA
	Sex	Male	0.015	0.024
	Maternal age		-0.016	0.190
	Maternal age^2		0.0001	0.016
August weight				
	Litter size	Singleton	0.000	NA
	Litter size	Twin	-3.441	0.145
	Sex	Female	0.000	NA
	Sex	Male	1.521	0.107
	Maternal age		1.742	0.093
	Maternal age^2		-0.134	0.008
	Capture age		0.066	0.008
Jaw length				
	Age at death		0.059	0.006
Metacarpal length				
	Age at death		0.009	0.006